

Supplementary Text

Both studies were conducted according to Declaration of Helsinki principles and International Committee on Harmonisation Good Clinical Practice guidelines. Governing ethical bodies at each study site approved the protocol, and all patients provided written informed consent before the conduct of study-specific procedures. An independent data monitoring committee regularly reviewed unblinded interim patient data for both studies.

Phase 2a methods

This was a multicenter, randomized, double-blind, placebo-controlled, parallel-group study in patients with active rheumatoid arthritis (RA) despite methotrexate (MTX) therapy (7.5-25 mg/week for ≥ 4 months). Approximately 90 patients were to be randomized (2:1) to toreforant 100 mg or matching placebo capsules given once daily for 12 weeks. Eligible patients also had a diagnosis of RA Functional Class I-III for ≥ 6 months prior to screening. Patients were allowed to continue stable doses of oral steroids (prednisone ≤ 10 mg/day, or equivalent corticosteroid) and nonsteroidal anti-inflammatory drugs (NSAIDs). Patients must have been treated with ≤ 3 antirheumatic biologic agents and ≤ 5 synthetic, conventional disease-modifying antirheumatic drugs (DMARDs) before screening.

The Phase 2a trial was terminated prematurely because of a serious adverse event (SAE) of secondary hemophagocytic lymphohistiocytosis (sHLH) with a fatal outcome in a patient receiving toreforant 100 mg/day. All efficacy results discussed are based on posthoc analyses that were defined to appropriately evaluate the treatment effects of toreforant, while reflecting the premature termination of the study.

Core clinical and laboratory evaluations comprised swollen joint count, tender joint count, the Physician's and Patient's Global Assessments of Disease Activity, Patient's Assessment of Pain, erythrocyte sedimentation rate (ESR), and C-reactive protein (CRP). Key clinical response assessments derived from these data included the 28-joint Disease Activity Score employing CRP (DAS28-CRP)¹ and the $\geq 20\%$ improvement in the American College of Rheumatology (ACR20) response criteria². The Health Assessment Questionnaire-Disability Index (HAQ-DI)³, the minimal clinically important improvement in which is considered 0.22 points⁴, also was administered.

The *post hoc* efficacy analyses conducted reflected the premature termination of the study and employed an analysis set comprising on-treatment observations (i.e., within 20 days of the last dose), with last-observation-carried-forward data imputation performed for the last value before treatment failure for four patients who received prohibited medications due to RA worsening during study participation. For such exploratory analyses, statistical significance should be considered as nominal.

Also in the *post hoc* efficacy analyses, baseline-adjusted analysis of covariance was employed to estimate treatment effects for change in DAS28-CRP from baseline, while the Cochran Mantel Haenszel (CMH) test was employed to determine statistically significant treatment differences in categorical response measures. Nominal statistical significance was assessed with respect to 2-sided testing at $p \leq 0.05$. Adverse events were to be coded using the current version of Medical Dictionary for Regulatory Activities (MedDRA) and summarized by system organ class, and by severity and relationship to study agent.

Supplementary Text (cont.)

Phase 2b methods

Phase 2b trial participants were adults with RA for ≥ 6 months, with a positive test for either anti-cyclic citrullinated peptide antibody or rheumatoid factor in serum, ≥ 6 tender and ≥ 6 swollen joints, and serum CRP ≥ 0.80 mg/dL. Patients must have received MTX (10-25 mg/week or 6-16 mg/week for Japanese patients) for ≥ 6 months, with a stable dose for ≥ 8 weeks. Other synthetic DMARDs were not permitted and must have been washed out. Patients were permitted to receive stable doses of oral corticosteroids (prednisone ≤ 10 mg/day or equivalent) and/or stable doses of NSAIDs or other analgesics. Patients who had ever received any approved or investigational biologic agent for a rheumatic indication were excluded.

Two hundred seventy-two patients were randomly (1:1:1:1) assigned to receive toreforant 3, 10, or 30 mg/day or placebo once daily for 24 weeks. The dose levels were selected to maintain trough concentrations greater than required for efficacy in a mouse collagen-induced arthritis model. Experience with histamine H₄-receptor (H₄R) antagonists in this model suggests that trough levels averaging 550 nM are the best predictors of efficacy. Accounting for the difference in the inhibitory constant (K_i) concentrations between the human and mouse receptors (8 nM and 300 nM, respectively), minimum concentration (C_{min}) values above 15 nM (6 ng/mL) were expected to be required for efficacy in humans. All doses studied (3-, 10-, 30- and 100-mg) should have been at the top of the dose-response curve and yielded comparable efficacy based on the mean C_{min} values being above 6 ng/mL.

The maximum length of study participation was to be 34 weeks, including a 6-week screening period, a 24-week placebo-controlled treatment period, and a 4-week follow-up period between the last dose and the last patient visit. Two database locks (DBLs) were planned (week 12 and week 28). At week 16, patients with $< 20\%$ improvement from baseline in both tender and swollen joint counts entered double-blind early escape. Such patients in the placebo, 3 mg/day, and 10 mg/day groups escaped to toreforant 30 mg/day, while no dose adjustments were made for patients already receiving 30 mg/day. The study was terminated after results deriving from the week-12 DBL (primary analysis) showed no evidence of efficacy with any of the three doses of toreforant evaluated.

Efficacy analyses (both from the week-12 and week-28 DBLs) were based on the modified intent-to-treat (ITT) population, i.e., randomized patients who received ≥ 1 dose of study agent. The primary efficacy endpoint was change from baseline in DAS28-CRP¹ at week 12. Secondary efficacy endpoints included ACR20 response² and change from baseline in the HAQ-DI scores³ at week 12 and week 24 and change from baseline in DAS28-CRP at week 24. Blood samples were collected for the measurement of plasma concentrations of toreforant and/or metabolites, and safety assessments included assessments of AEs, clinical laboratory determination, vital signs, and electrocardiograms (ECGs).

Sample size and power were evaluated by estimating the probabilities of establishing a dose-response relationship using the Multiple Comparison Procedures with modeling techniques method. To preserve the overall Type I error rate, the significance level of 0.05 with 2-sided test was used. Response was defined as the improvement in the DAS28-CRP between week 0 and week 12. Seventy patients per arm were planned to provide $> 90\%$ power with a standard deviation (SD) of 1.4 for detecting a dose-response signal in change in DAS28-CRP for all models. In addition, 70 patients per group would to provide $> 80\%$ power in detecting a

Supplementary Text (cont.)

difference of $\geq 25\%$ in ACR20 response at week 12 for a toreforant dose group compared with the placebo group. For all secondary efficacy endpoints, p values for treatment group comparisons were derived from a two-sided CMH test for categorical endpoints or a two-sided analysis of variance/covariance for continuous endpoints.

Improvement from baseline in the DAS28-CRP at week 12, the primary efficacy endpoint, was used as the response variable in the dose-response analysis. In addition to the dose-response curve, the changes from baseline in DAS28-CRP at week 12 were subjected to pairwise comparisons between each toreforant dose group and the placebo group based on an analysis of variance model yielding nominal p-values.

Toreforant concentration-time data from all patients was to be pooled and subjected to population pharmacokinetic analysis using nonlinear mixed effects modeling. All safety analyses were to be performed on the safety analysis set, which included all randomized patients who received ≥ 1 dose of the study agent. Adverse events were coded using the current version of MedDRA and summarized by system organ class, and by severity and relationship to study agent.

Phase 2a results

The toreforant Phase 2a trial was conducted at 26 sites in nine countries (Russia-43% of patients, Czech Republic-26%, South Korea-13%, Poland-5%, United Kingdom-3%, Netherlands-3%, Taiwan-3%, Spain-2%, and Ireland-1%) from 12/9/2009-11/17/2010. The study was terminated prematurely because of an SAE of sHLH with a fatal outcome in a patient receiving toreforant 100 mg/day. A 38-year-old woman died unexpectedly 21 days after starting toreforant 100 mg/day. The patient's rapid deterioration, her fatal outcome, and the initially unclear cause/relationship to study drug led to immediate study termination. Subsequently, an extensive investigation including consultation with external subject-matter clinical experts, indicated the most likely cause of death was sHLH, a rare immune activation syndrome typically occurring alongside autoimmune disease. Given sHLH pathogenesis, the known effects of H₄R-antagonism, the preclinical toxicological toreforant profile, and the fact that the patient carried a heterozygous mutation of perforin gene, the causal relationship between toreforant and sHLH was considered unlikely. Additional SAEs reported for this patient, i.e., disseminated intravascular coagulation, pancytopenia, hepatitis acute, metabolic acidosis, acute prerenal failure, azotemia, and hypotension, were clinical manifestations of sHLH. The other SAEs reported for additional patients receiving toreforant 100 mg/day were RA, arthralgia, yersinia infection, and spinal compression fracture.

No meaningful changes from baseline in corrected QT interval were observed when measured at approximate time of maximum concentrations at week 2 (toreforant and placebo mean changes of 0.73 and 0.24 msec, respectively) or at week 12 (1.77 and -5.12 msec, respectively).

Phase 2b trial additional results

The toreforant Phase 2b study was conducted in Japan (11.4% of patients), Taiwan (0.4%) Thailand (5.9%), Czech Republic (4.0%), Hungary (9.2%), Latvia (4.8%), Poland (8.5%), Romania (0.7%), Russia (15.4%), Ukraine (7.0%), Argentina (1.5%), Chile (4.8%), Colombia (7.4%), Mexico (15.4%), and the United States (3.7%).

Supplementary Text (cont.)

Additional secondary efficacy outcomes assessed included the Clinical Disease Activity Index (CDAI)⁵ and the Simplified Disease Activity Index (SDAI)⁶. The 36-item short-form medical survey (SF-36) physical and mental component summary (PCS and MCS) scores⁷ was used to assess quality of life.

At week 12, no significant differences in CDAI or SDAI median changes from baseline were observed between the combined toreforant (-15.20 and -15.64, respectively) and placebo (-16.70 and -16.55, respectively) groups ($p=0.175$ and 0.202 , respectively). Additionally, no apparent differences between the combined toreforant and placebo group were observed for median SF-36 PCS (4.20 vs. 2.98, respectively) or MCS (2.55 vs. 2.66, respectively) scores at week 24.

REFERENCES

1. Prevoo ML, van' t Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight joint counts: development of validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995;38:44-8.
2. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;38:727-35.
3. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. *J Rheumatol* 1982;9:789-93.
4. Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol* 1993;20:557-60.
5. Aletaha D, Nell VP, Stamm T, Uffmann M, Pflugbeil S, Machold K, et al. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. *Arthritis Res Ther* 2005;7:R796-R806.
6. Smolen JS, Breedveld FC, Schiff MH, Kalden JR, Emery P, Eberl G, et al. A simplified disease activity index for rheumatoid arthritis for use in clinical practice. *Rheumatology* 2003;42:244-57.
7. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.

Supplementary Table 1

Table S1. Summary of patient disposition through week 12 in the Phase 2a study or through week 24 in the Phase 2b study.

	Phase 2a			Phase 2b			
	Toreforant		Placebo	-----Toreforant-----			Combined
	Placebo	100 mg		3 mg	10 mg	30 mg	
Randomized patients, N	28	58	68	68	68	68	204
Modified intent-to-treat population ¹ , N	28	58	68	67	68	68	203
Patients completed, n (%)	12 (43)	24 (41)	57 (83.8)	50 (73.5)	57 (83.8)	53 (77.9)	160 (78.4)
Patients prematurely discontinued study agent, n (%)	16 (57)	34 (59)	11 (16.2)	18 (26.5)	11 (16.2)	15 (22.1)	44 (21.6)
Reason for discontinuation							
-Adverse Event	3 (11)	4 (7)	5 (7.4)	7 (10.3)	3 (4.4)	3 (4.4)	13 (6.4)
-Withdrawal of consent	0	1 (2)	0	2 (2.9)	0	2 (2.9)	4 (2.0)
-Lost to follow-up	0	0	0	0	0	1 (1.5)	1 (0.5)
-Physician decision	--	--	0	0	0	1 (1.5)	1 (0.5)
-Lack of efficacy	--	1 (2)	2 (2.9)	3 (4.4)	3 (4.4)	5 (7.4)	11 (5.4)
-Other	--	--	4 (5.9)	6 (8.8)	5 (7.4)	3 (4.4)	14 (6.9)
-Sponsor early discontinuation	13 (46)	28 (48)	4 (5.9)	5 (7.4)	4 (5.9)	3 (4.4)	12 (5.9)

¹ Defined as all randomized patients who received ≥ 1 dose of study agent.

Supplementary Table 2

Table S2. Summary of baseline patient demographics in the Phase 2a and Phase 2b studies.

	Phase 2a			Phase 2b			
	Placebo	Toreforant		-----Toreforant-----			
		100 mg	Placebo	3 mg	10 mg	30 mg	Combined
Randomized patients, N	28	58	68	68	68	68	204
Age (years), mean (SD)	54.6 (10.0)	50.2 (11.5)	51.5 (11.4)	52.8 (11.4)	49.5 (13.2)	53.8 (12.6)	52.0 (12.5)
Female, n (%)	20 (71)	48 (83)	57 (83.8%)	60 (88.2%)	52 (76.5%)	51 (75.0%)	163 (79.9%)
Duration of RA (years), mean (SD)	6.6 (5.6)	6.9 (5.9)	5.9 (5.2)	8.9 (7.2)	7.0 (6.2)	8.4 (7.6)	8.1 (7.1)
HAQ-DI score (0-3), mean (SD)	1.6 (0.5)	1.6 (0.6)	1.6 (0.7)	1.6 (0.7)	1.5 (0.7)	1.7 (0.7)	1.6 (0.7)
CRP (mg/dL), mean (SD)	2.4 (2.7)	2.2 (2.0)	2.1 (1.7)	2.6 (2.6)	2.0 (2.0)	2.4 (2.2)	2.3 (2.3)

CRP: C-reactive protein; HAQ-DI: Health Assessment Questionnaire-Disability Index; RA: rheumatoid arthritis.

Supplementary Figure 1

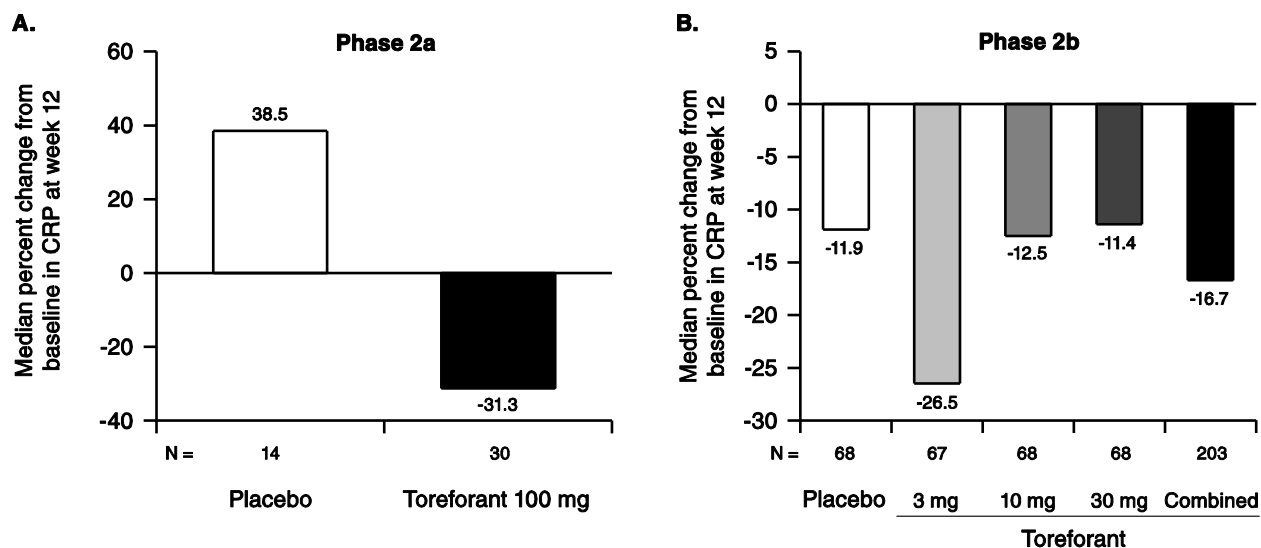


Figure S1. Median percent changes from baseline in CRP levels at Week 12 in the phase IIa study (observed cases on treatment with LOCF of last visit before treatment failure; A) and in the phase IIb study (mITT population; B). CRP: C-reactive protein; LOCF: last observation carried forward; mITT: modified intent-to-treat.

Supplementary Figure 2 (see below)

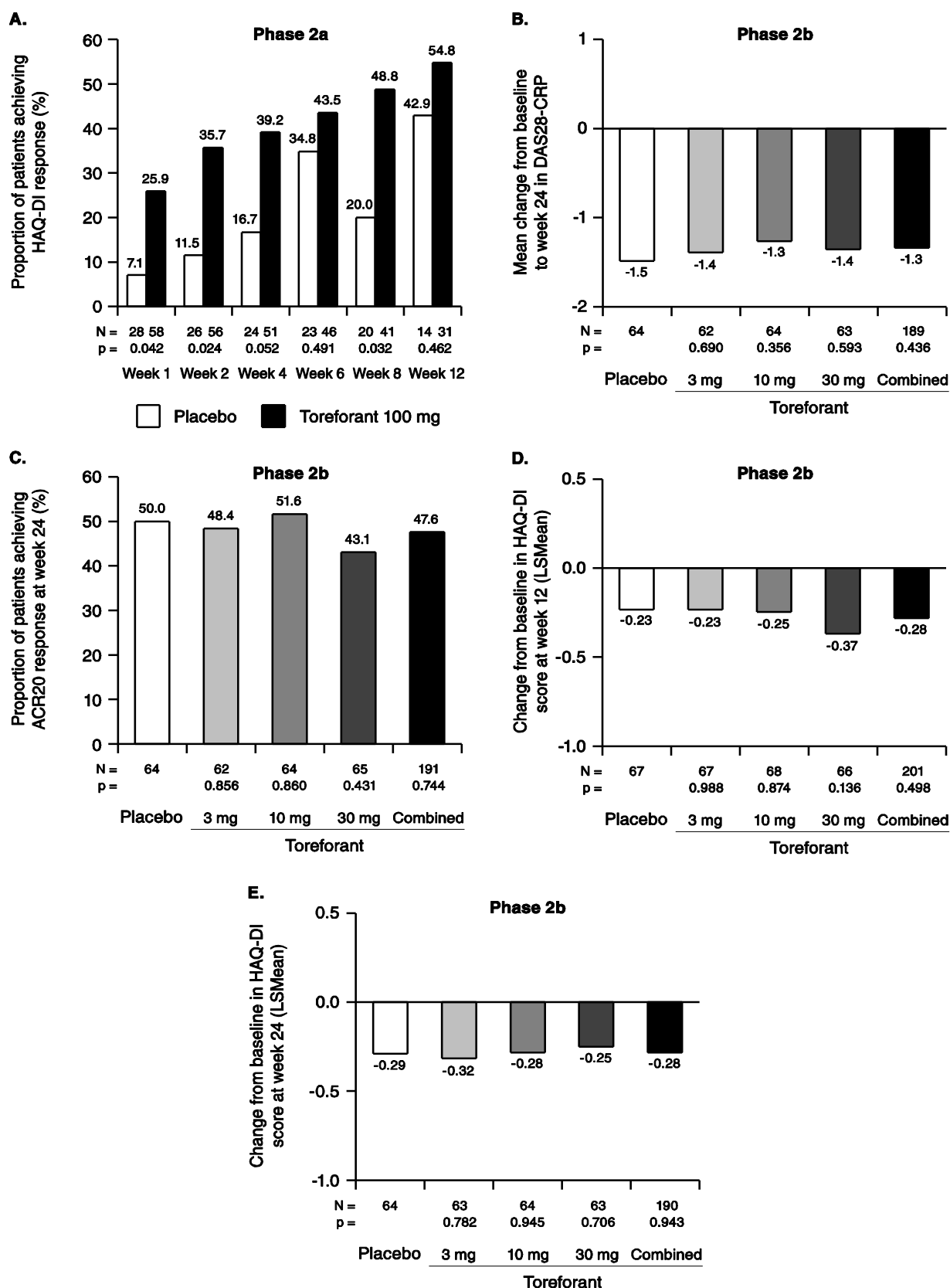


Figure S2. Proportions of patients achieving HAQ-DI response (A), defined as improvement ≥ 0.30 , through Week 12 in the phase IIa study (observed cases on treatment with LOCF of last visit before treatment failure), as well as mean changes from baseline in DAS28-CRP at Week 24 (B), proportions of patients achieving ACR20 response at Week 24 (C), and mean changes from baseline in HAQ-DI score at Week 12 (D) and Week 24 (E) in the phase IIb study (mITT population). ACR20: American College of Rheumatology 20% improvement; DAS28-CRP: 28-joint Disease Activity Score incorporating C-reactive protein; HAQ-DI: Health Assessment Questionnaire-Disability Index; LOCF: last observation carried forward; LSMean: Least Square Mean; mITT: modified intent-to-treat.